**Assessment of current disease burden and unmet needs in MDD**

**SCAN-2030 Work Package 1 [WP1] Analysis Plan**

## 1. OBJECTIVE

To assess the ten-year current disease burden, unmet care needs and economic burden related to major depressive disorder ("depression" thereafter) in Hong Kong.

## 2. RATIONALE

Information on disease burden, unmet needs and economic burden will form the foundation for innovative medicine decision-making for both supply and demand. We will demonstrate how real-world data could be used to understand current care needs as potential tools to guide health policy and marketing decisions.

## 3. DATA SOURCE

We will utilise Clinical Data Analysis and Reporting System (CDARS), a territory-wide electronic medical record (EMR) database managed by the Hospital Authority in Hong Kong. Real-time records in patient demographics, dates of registered death, dates of hospitalization and service attendance, all-cause diagnoses, prescriptions, procedures and laboratory tests across inpatient, outpatient and emergency settings are centralized for audit and research purposes, and de-identified to protect patient confidentiality.

## 4. STUDY POPULATION

Patients with clinical diagnosis of depression between 1 January 2012 and 31 December 2022 will be identified from the EMR database using International Classification of Diseases, 9$^{th}$ Revision, Clinical Modification (ICD-9-CM) diagnostic codes (296.2, 296.3, 300.4 and 311). The patients will be divided into 11 cohorts based on year of diagnosis. The cohort follow-up period will be from the date of cohort identification until death or study end date (31 December 2022).

## 5. STUDY OUTCOMES

### 5.1 Prevalence of active patients using Hospital Authority services

"Prevalent patients" mean the current cases living with the disease, no matter if they are newly diagnosed or pre-existing cases in the current time point. "Prevalence", or technically "period prevalence", refers to the number of current cases divided by the population or sample during

a specified period. In this part, we will identify the **annual unique number of patients whoever received a clinical diagnosis of depression** in each year from 2012 to 2022, no matter if they were newly diagnosed in the current year or the years before. We will also calculate the **annual crude and age-standardardized prevalence** by dividing the number of cases by the total mid-year local population in the corresponding year*. For both types of outcome (annual case number and prevalence per 10,000 persons), we will describe and illustrate the trend of overall, age-specific and sex-specific figures from 2012 to 2022. Finally, we will calculate the **ten-year period prevalence** by dividing the total number of unique prevalent patients from 2012 to 2022 (nominator) by the total mid-year population of 2017.
*Source: https://www.censtatd.gov.hk/en/web_table.html?id=110-01002#

**Tips for analysis:** Combine the "all diagnosis" data, which recorded the all-cause diagnoses from 1993 to 2022 of each cohort. Subset the data to include only records with ICD-9-CM codes of 296.2, 296.3, 300.4, 311. To find the number of patients whoever had a depression diagnosis in, for example 2014, simply subset the records with reference dates fallen into the year of 2014. Remove the duplicated reference keys then count the number of reference keys. Please note that this would be an **underestimated count of true prevalence** because it only accounts for the patients who actively sought consultation in the public healthcare.

### 5.2 Incidence & newly diagnosed patients

"Incident patients" are the new cases of the disease. "Incidence", or technically incidence proportion, refers to the number of new cases during a specific period divided by population at the start of the interval. In this part, we will identify the **annual number of patients newly diagnosed with depression** in each year from 2012 to 2022. We will also calculate the **annual crude and age-standardized incidence** by dividing the number of new cases by the total local population at the start of the corresponding year. The trend of both outcomes will be described and illustrated at overall, age-specific and sex-specific settings.

**Tips for analysis:** Within the list of prevalent patients obtained from the section 5.1, check whether they had previous depression diagnosis from 1993, which is the year in which the database was available, to the year before diagnosis. Patients with no history before could be concluded as new cases in that calendar year. ICD-9 code 296.3 (Major Depressive Disorder, Recurrent Episode) applies to patients with multiple episodes of major depression. This means they cannot be newly diagnosed with this code, as it assumes at least one prior depressive

episode. Therefore, patients newly diagnosed with ICD-9 code 296.3 should be excluded from the incident cohort.

### 5.3 All-cause mortality

We will follow up the incident cohorts, i.e. patients who were newly diagnosed in the defined year, starting from the earliest date of confirmatory depression diagnosis (index date) to death or the end of study (31 December 2022). We will report the **trend of all-cause mortality** of each cohort, with subgroup analyses stratified by age groups and sex. We will also report the **annual number of deaths, annual mortality rates** and **10-year mortality rates** until 2022. Mortality rates refer to the total number of deaths in the specified period divided by the number of cases who were at risk of dying for that period. Similar analysis may be performed for all cohorts between 2013 and 2022. We will also combine all incident cohorts and perform survival analysis using Kaplan-Meier plot with follow-up until December 2022.

### 5.4 One-year costs of care under Hospital Authority

We will report the annual cost of all-cause care from 2012 to 2022. Based on the 11 prevalent cohorts identified in section 5.1, we will follow up on the patterns of healthcare resource utilization from the index date to death or the annual window cut-off date for each cohort. Taking the 2012 prevalent cohort as an example, the follow-up period of new cases in 2012 will be from the first date of diagnosis to death or 31 December 2012, and the follow-up period of pre-existing cases in 2012 will be from 1 January 2012 to death or 31 December 2012. There will be in total 11 one-year costs which trend can be illustrated.

During the follow-up period of each cohort, we will identify the total number of attendance episodes in the outpatient settings and the total lengths of stay (LOS) in the accident & emergency (A&E) and inpatient settings in a service-type-specific manner. The total episodes or LOS in the 15 service types will be multiplied by the service-specific unit costs (https://www.ha.org.hk/visitor/ha_visitor_index.asp?Content_ID=10045&Lang=ENG) charged as non-eligible persons by the Hospital Authority. There will be 11 aggregated costs which will then be used to plot the graph.

#### 5.4.1 Inpatient by-ward bed-days

Four ward-specific LOS are calculated as follows:

General = [LOS of Ward Care Type Acute General - Acute] + [LOS of Ward Care Type Convalescent / Rehabilitation / Infirmary]
Psychiatric = [LOS of Ward Care Type Psychiatry / Mentally Handicapped]
High dependency = [LOS of Ward Care Type Acute General - High Dependency]
ICU = [LOS of Ward Care Type Acute General - Intensive Care]

### 5.4.2 Outpatient service-specific episodes

Outpatient service classification of each record is based on the reference coding table used in previous studies. Please note that not all records will be counted because some services types are not associated with a specific cost (e.g. health education).

### 5.4.3 All-cause and psychiatric cost of care

After obtaining the inpatient, outpatient and A&E costs, we will be able to sum them up to obtain the overall all-cause cost of care. Psychiatric-related cost of care will be equal to costs involved in inpatient psychiatric ward + psychiatric day hospital + SOPC(psychiatric) + community(psychiatric).

## 5.5 Unmet Needs for innovative medicines

Using the prescription data, we will visualize the number of patients in each incident cohort (2012-2021) who developed treatment-resistant depression during the follow-up until 2022. Treatment-resistant depression is defined as at least two trials/switches of antidepressant regimens with adequate duration and dosage, with the presence of the third regimen to confirm the failure of the previous two lines of treatment. To be more specific, one regimen is defined as the same choice of antidepressant or combination regimen of at least 28 days with gaps no longer than 14 days. The date of turning into a treatment-resistant patient is defined as the date of commencing the third regimen. Patients who do not meet the criteria for TRD throughout the entire follow-up period (until death or end of study period) will be classified as non-TRD. We will plot the cumulative proportion of patients who develop TRD by year in all cohorts.

## 5.6 Treatment Trajectory

Based on treatment records, we used a Sankey diagram to visualize the progression of patients in each incident cohort (2012-2021) from their initial treatment to the development of treatment resistance. The diagram illustrated the proportion of patients on various treatment regimens at three distinct stages, showing the flow of treatments over the follow-up period.

The follow-up began at diagnosis and continued until either death or the end of the observation period (12-31-2022). The first treatment regimen consisted of antidepressant monotherapy, with antidepressants categorized by their pharmacological mechanisms. For the second and third regimens, patients could receive either different antidepressant monotherapies or combination therapies, which could include augmentation with second-generation antipsychotics or mood stabilizers. However, due to data limitations, behavioural therapies and psychotherapies were not included in the analysis.